Stochastic Constraints for Fast Image Correspondence Search with Uncertain Terrain Model

Michael Veth, *Student Member, IEEE*, John Raquet, *Member, IEEE*, Meir Pachter, *Fellow, IEEE*,
Air Force Institute of Technology

***Abstract*— The navigation state (position, velocity, and attitude) can be determined using optical measurements from an imaging sensor pointed toward the ground. Extracting navigation information from an image sequence depends on tracking the location of stationary objects in multiple images, which is generally termed the correspondence problem. This is an active area of research and many algorithms exist which attempt to solve this problem by identifying a unique feature in one image and then searching subsequent images for a feature match. In general, the correspondence problem is plagued by feature ambiguity, temporal feature changes, and occlusions which are difficult for a computer to address. Constraining the correspondence search to a subset of the image plane has the dual advantage of increasing robustness by limiting false matches and improving search speed. A number of ad-hoc methods to constrain the correspondence search have been proposed in the literature.**

In this paper, a rigorous stochastic projection method is developed which constrains the correspondence search space by incorporating *a priori* knowledge of the aircraft navigation state using inertial measurements and a statistical terrain model. The stochastic projection algorithm is verified using Monte Carlo simulation and flight data. The constrained correspondence search area is shown to accurately predict the pixel location of a feature with an arbitrary level of confidence, thus promising improved speed and robustness of conventional algorithms.

I. Introduction

IT is well-known that optical measurements provide excellent navigation information, when interpreted properly. Optical navigation is not new. Pilotage is the oldest and most natively familiar form of navigation to humans and other animals. For centuries, navigators have utilized mechanical instruments such as astrolabes, sextants, and driftmeters [12] to make precision observations of the sky and ground in order to determine their position, velocity, and attitude.

The difficulty in using optical measurements for autonomous navigation, that is, without human intervention, has always been in the interpretation of the image, a difficulty shared with Automatic Target Recognition (ATR). Indeed, when celestial observations are used, the ATR problem in this structured environment is tractable and automatic star trackers are widely used for space navigation and ICBM guidance. When ground images are to be used, the difficulties associated with image interpretation are paramount. At the same time, the problems associated with the use of optical measurements for navigation are somewhat easier than ATR. Moreover, recent developments in feature tracking algorithms, miniaturization, and reduction in cost of inertial sensors and optical imagers, aided by the continuing improvement in microprocessor technology, motivates us to consider using inertial measurements to aid the task of feature tracking in image sequences.

The methods are typically classified as either feature-based or optic flow-based, depending on how the image correspondence problem is addressed. Feature-based methods determine correspondence for "landmarks" in the scene over multiple frames, while optic flow-based methods typically determine correspondence for a whole portion of the image between frames. A good reference on image correspondence is [7]. Optic flow methods have been proposed in the literature generally for elementary motion detection, focusing on determining relative velocity, angular rates, or for obstacle avoidance [4].

Feature tracking-based navigation methods have been proposed both for fixed-mount imaging sensors or gimbal mounted detectors which "stare" at the target of interest, similar to the gimballed infrared seeker on heat-seeking, air-to-air missiles. Many feature tracking-based navigation methods exploit knowledge (either *a priori*, through binocular stereopsis, or by exploiting terrain homography) of the target location and solve the inverse trajectory projection problem [1], [10]. If no *a priori* knowledge of the scene is provided, egomotion estimation is completely correlated with estimating the scene. This is referred as the structure from motion (SFM) problem. A theoretical development of the geometry of fixed-target tracking, with no *a priori* knowledge is provided in [11]. An online (Extended Kalman Filter-based) method for calculating a trajectory by tracking features at an unknown location on the Earth's surface, provided the topography is known is given in [3]. Finally, navigation-grade inertial sensors and terrain images collected on a T-38 "Talon" are processed and the potential benefits of optical-aided inertial sensors are experimentally demonstrated in [14].

Many methods for solving the correspondence problem have

M. Veth, Major, U.S. Air Force, Air Force Institute of Technology, 2950 Hobson Way, Wright-Patterson Air Force Base, OH 45433, USA michael.veth@afit.edu

J. Raquet, Associate Professor, Department of Electrical and Computer Engineering, Air Force Institute of Technology, 2950 Hobson Way, Wright-Patterson Air Force Base, OH 45433, USA john.raquet@afit.edu

M. Pachter, Professor, Department of Electrical and Computer Engineering, Air Force Institute of Technology, 2950 Hobson Way, Wright-Patterson Air Force Base, OH 45433, USA meir.pachter@afit.edu

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2007		**00-00-2007 to 00-00-2007**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Stochastic Constraints for Fast Image Correspondence Search with Uncertain Terrain Model	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Air Force Institute of Technology,2950 Hobson Way,Wright Patterson AFB,OH,45433-7765**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**8**	

been proposed in the computer vision literature. A popular algorithm is the Lucas-Kanade feature tracker [6], which relies on the premise of the invariance of the intensity field between images. It uses a template correlation algorithm to minimize the sum of squared differences (SSD) between image intensities. The algorithm typically assumes a linear ($x - y$ plane) motion model, but can be extended to optimize over affine or bilinear transformations. Other feature correspondence algorithms have been proposed which are invariant to rotations, scaling or both. (e.g., [5]) More robust feature tracking algorithms are typically computationally expensive and a designer must trade tracking robustness and accuracy for real-time performance.

This paper proposes an approach to optimize the feature tracking problem by exploiting navigation information, derived from six degree-of-freedom inertial measurements, and prior terrain information, to constrain the correspondence search space and aid the attendant optimization algorithm. The theory is developed for a kinematic motion model with inertial sensors.

The paper is organized as follows. Section II explores current approaches for constraining correspondence searches and discusses the strengths and weaknesses of such. Section III poses the statistical projection problem in the most general terms. Reasonable assumptions are proposed which make the general problem tractable for use with an Extended Kalman Filter algorithm. In Section IV, the mathematical model used to describe the navigation state and navigation state uncertainty is presented. This includes definition of reference frames, navigation dynamics, perturbation model, and defines the initial conditions. Section V builds upon the mathematical model to derive the stochastic projection method. The resulting equations allow the user to predict the pixel location and uncertainty of a feature between two images. The stochastic projection method is validated in Section VI using Monte Carlo simulation and flight data. Finally, conclusions are drawn regarding the performance of the method in Section VII.

II. Current Correspondence Constraint Approaches

Exploiting inertial measurements to constrain the correspondence search has been proposed in the literature. In this section, two methods which exploit inertial measurements are discussed.

Bhanu and Roberts [2] utilize inertial measurements to compensate for rotation between images and to predict the focus of expansion in the second image. Once the second image is derotated and the focus of expansion is established, the correspondence between points of interest is calculated using goodness-of-fit metrics. One relevant metric is the correspondence search constraint placed on each point. This constraint ensures each interest point lies in a cone-shaped region, with apex at the focus of expansion, bisected by the line joining the focus of expansion and the interest point in the camera frame at the first image time. While this constraint is not statistically rigorous, it does show the value of using inertial measurements to aid the correspondence problem.

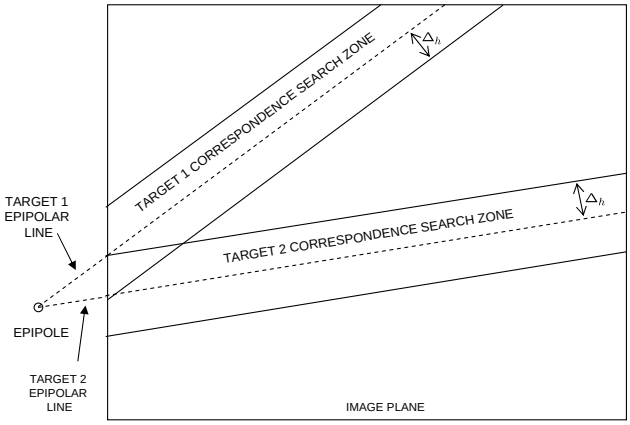

Fig. 1. Correspondence search constraint using epipolar lines. Given a projection of an arbitrary point in an initial image, combined with knowledge of the translation and rotation to a second image, the correspondence search can be constrained to an area near the epipolar line. Note the epipole can be located outside of the image plane, as shown in this example.

Strelow also incorporates inertial measurements to constrain the correspondence search between image frames [15]. This constraint on the image search space is a similar concept to the field of expansion method proposed by Bhanu; however, Strelow generalizes the approach by exploiting epipolar geometry.

The projection of an arbitrary point in an image is described by an epipolar line in a second image. All epipolar lines in an image converge at the projection of the focus of the complimentary image. Combining knowledge of the translation and rotation between images and the pixel location of a candidate target in the first image, a correspondence search can then be constrained to an area "near" the epipolar line. This approach is illustrated in Fig. 1. Strelow's method of using inertial measurements to constrain the correspondence search along an epipolar line is ad-hoc, since the search space is not defined statistically.

In the next Section, the correspondence problem is described using a stochastic model. This model is then used to determine a statistically-rigorous correspondence search area.

III. General Problem Formulation

The general problem is described as follows. Given a pixel location of a specified landmark at time t_i, predict the probability density function of the pixel location of the same landmark at time t_{i+1}. Prior information regarding the vehicle navigation state, terrain statistics, and the dynamics of the vehicle and landmark are exploited.

Mathematically, the pixel location, $\mathbf{z}(t_i)$, corresponding to a landmark at location $\mathbf{y}(t_i)$ in the scene, is governed by the nonlinear projection function

$$\mathbf{z}(t_i) = \mathbf{h}\left[\mathbf{x}(t_i), \mathbf{y}(t_i), t_i\right] \tag{1}$$

where $\mathbf{x}(t_i)$ represents the navigation state at the time of the measurement.

The vehicle and landmark dynamics are modeled by the following non-linear Itô stochastic differential equations in

white noise notation,

$$\dot{\mathbf{x}}=\mathbf{f}\left[\mathbf{x}(t),\mathbf{u}(t),t\right]+\mathbf{G}_x\left[\mathbf{x}(t),t\right]\mathbf{w}_x(t) \quad (2)$$

$$\dot{\mathbf{y}}=\mathbf{r}\left[\mathbf{y}(t),t\right]+\mathbf{G}_y\left[\mathbf{y}(t),t\right]\mathbf{w}_y(t) \quad (3)$$

where $\mathbf{u}(t)$ is a known input function, and $\mathbf{w}_x(t)$ and $\mathbf{w}_y(t)$ are white noise processes.

A theoretical formulation exists for this general problem, however two issues make this solution intractable. First, the measurement observation function is ill-posed (i.e., a unique inverse does not exist). Second, propagating the conditional probability density function in time requires solving the forward Kolmogorov (i.e., Fokker-Planck) second-order partial differential equation for an infinite number of moments [9].

To make the problem tractable, the following reasonable assumptions are made:

- The prior knowledge of the navigation and target state can be adequately described as a multivariate Gaussian distribution.
- Additive measurement noise is zero-mean, Gaussian, and white.
- Stochastic process noise is zero-mean, Gaussian, and white.
- The nonlinear state dynamics and measurement equations can be adequately modeled using perturbation techniques.

Although not required for tractability, additional assumptions are made to simplify the development and clarify the underlying concepts. First, the landmark is assumed to be stationary with respect to the surface of the Earth (i.e., $\mathbf{r}[\mathbf{y}(t),t] = 0$.) Second, the camera is rigidly mounted to the vehicle with known alignment and calibration. Third, the terrain is described by a statistical elevation model.

In the next section, the relevant reference frames and the vehicle dynamics are defined.

IV. Mathematical Model

A. Reference Frames

In this paper, three reference frames are used. Variables expressed in a specific reference frame are indicated using superscript notation. The Earth-Centered Earth-Fixed (ECEF, or *e* frame) is a Cartesian system with the origin at the Earth's center, the $\hat{x}^e$ axis pointing toward the intersection of the equator and the prime (Greenwich) meridian, the $\hat{z}^e$ axis extending through the North pole, and the $\hat{y}^e$ axis is the orthogonal compliment (in this paper, a carat symbol, $\hat{}$, denotes a unit vector). The navigation state is expressed in the *e* frame.

The vehicle body frame (or *b* frame) is a Cartesian system with origin at the vehicle center of gravity, the $\hat{x}^b$ axis extending through the vehicle's nose, the $\hat{y}^b$ axis extending through the vehicle's right side, and the $\hat{z}^b$ axis points orthogonally out the bottom of the vehicle. The inertial measurements are expressed in the *b* frame.

The camera frame (or *c* frame) is a Cartesian system with origin at the center of the camera image plane, the $\hat{x}^c$ axis is parallel to the camera image plane and defined as "camera up", the $\hat{y}^c$ axis is parallel to the camera image plane and defined as "camera right", and the $\hat{z}^c$ axis points out of the camera aperture, orthogonal to the image plane.

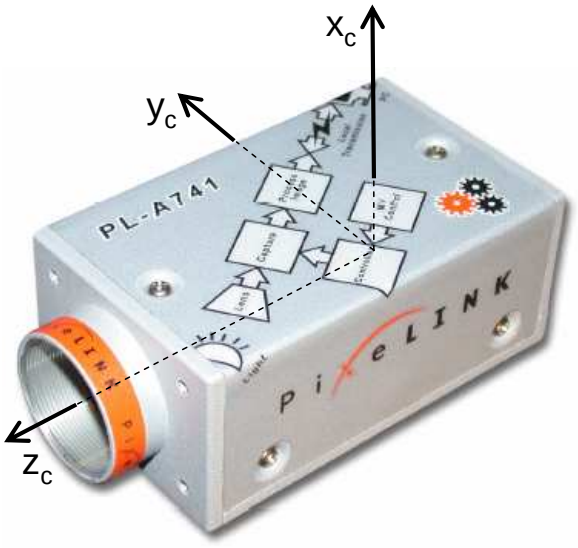

Fig. 2. Camera frame illustration. The camera reference frame originates at the center of the focal plane.

B. Vehicle State and Dynamics

The vehicle state of interest consists of position ($\mathbf{p}^e$), velocity ($\mathbf{v}^e$), and direction cosine matrix of the body to ECEF frame ($\mathbf{C}_b^e$). From [16], the vehicle state kinematics are

$$\dot{\mathbf{p}}^e = \mathbf{v}^e \quad (4)$$

$$\dot{\mathbf{v}}^e = \mathbf{C}_b^e\mathbf{f}^b - 2\boldsymbol{\Omega}_{\boldsymbol{ie}}^{\boldsymbol{e}}\mathbf{v}^e + \mathbf{g}^e \quad (5)$$

$$\dot{\mathbf{C}}_b^e = \mathbf{C}_b^e\boldsymbol{\Omega}_{ib}^b - \boldsymbol{\Omega}_{ie}^e\mathbf{C}_b^e \quad (6)$$

where $\mathbf{f}^b$ is the specific force vector measured by the accelerometers, $\boldsymbol{\Omega}_{ie}^e$ is the Earth's sidereal angular rate vector in skew-symmetric form, $\mathbf{g}^e$ is the gravitational acceleration vector, and $\boldsymbol{\Omega}_{ib}^b$ is the angular rate of the vehicle relative to the inertial frame in skew-symmetric form and measured by the gyroscopes.

C. Perturbation Model

The navigation errors are defined as differences from a nominal trajectory and are represented as a position error ($\delta\mathbf{p}^e$), a velocity error ($\delta\mathbf{v}^e$), and an attitude error ($\boldsymbol{\epsilon}$) vector, defined as:

$$\tilde{\mathbf{p}}^e(t) = \mathbf{p}^e(t) + \delta\mathbf{p}^e(t) \quad (7)$$

$$\tilde{\mathbf{v}}^e(t) = \mathbf{v}^e(t) + \delta\mathbf{v}^e(t) \quad (8)$$

$$\tilde{\mathbf{C}}_b^e(t) = [\mathbf{I}_3 - (\boldsymbol{\epsilon}(t)\times)]\mathbf{C}_b^e(t) \quad (9)$$

where the tilde represents a nominal parameter. The error state is modeled as a zero-mean Gaussian random vector

$$\delta\mathbf{x}(t) = \begin{bmatrix} \delta\mathbf{p}^e(t) \\ \delta\mathbf{v}^e(t) \\ \boldsymbol{\epsilon}(t) \end{bmatrix} \quad (10)$$

with covariance defined as

$$E[\delta\mathbf{x}(t)\delta\mathbf{x}^T(t)] = \mathbf{P}_{xx}(t) \quad (11)$$

where $E[\cdot]$ is the expectation operator.

Using perturbation techniques, the dynamics of the navigation error states are modeled as a linear stochastic differential equation [8]

$$\delta\dot{\mathbf{x}}(t) = \mathbf{F}(t)\delta\mathbf{x}(t) + \mathbf{G}_x[\tilde{\mathbf{x}}^e(t), t]\mathbf{w}_x(t) \tag{12}$$

where $\mathbf{w}_x(t)$ is a zero-mean, white Gaussian noise process with covariance kernel

$$E[\mathbf{w}_x(t)\mathbf{w}_x^T(t+\tau)] = \mathbf{Q}_x(t)\boldsymbol{\delta}(\tau). \tag{13}$$

D. Initial Conditions

At the time of the first image, t_i, the navigation error state is a zero-mean Gaussian random variable with covariance, $P_{xx}(t_i)$. The terrain elevation, h, is a random variable with mean, $\tilde{h}$, and variance, σ_h^2. The terrain elevation errors are assumed to be independent of the navigation errors.

V. Stochastic Projection Theory

The theory is divided into three sections: estimating the initial landmark position and covariance based on the pixel location of the feature selected in the first image, using inertial measurements to propagate this augmented state to the time of the second image, and projecting this landmark position state onto the second image as a probability density function in pixel coordinates. In simpler terms, these equations allow us to "predict" where a stationary feature should appear in subsequent images, thus providing a statistical measure to constrain our search space within the image.

A. Landmark Error Statistics

The landmark position corresponding to a pixel location is a non-linear function of the navigation state, pixel location, $\mathbf{z}(t_i)$, terrain elevation, h, camera to body direction cosine matrix, $\mathbf{C}_c^b$, and homogeneous camera projection matrix, $\mathbf{\Pi}$ (see [7] for a description):

$$\mathbf{y}^e = \mathbf{g}\left[\mathbf{p}^e(t_i), \mathbf{C}_b^e(t_i), \mathbf{z}(t_i), h, \mathbf{C}_c^b, \mathbf{\Pi}\right] \tag{14}$$

The pixel location measurement at time t_i is a non-linear function of the navigation state, landmark position, and camera parameters:

$$\begin{aligned}\tilde{\mathbf{z}}(t_i) &= \mathbf{h}[\mathbf{p}^e(t_i), \mathbf{C}_b^e(t_i), \mathbf{y}^e(t_i), \mathbf{C}_c^b, \mathbf{\Pi}] \\ &\quad +\mathbf{v}(t_i)\end{aligned} \tag{15}$$

where $\mathbf{v}(t_i)$ is a zero-mean, additive white Gaussian noise process with:

$$E[\mathbf{v}(t_i)\mathbf{v}(t_j)] = \begin{cases} \mathbf{R}(t_i) & t_i = t_j \\ \mathbf{0} & t_i \neq t_j \end{cases} \tag{16}$$

Similarly to the navigation state, the calculated landmark position, $\tilde{\mathbf{y}}^e$, is also modeled as a perturbation about the true position:

$$\tilde{\mathbf{y}}^e = \mathbf{y}^e + \delta\mathbf{y}^e \tag{17}$$

and is a function of the calculated trajectory

$$\tilde{\mathbf{y}}^e = \mathbf{g}\left[\tilde{\mathbf{p}}^e(t_i), \tilde{\mathbf{C}}_b^e(t_i), \tilde{\mathbf{z}}(t_i), \mathbf{C}_c^b, \mathbf{\Pi}\right] \tag{18}$$

Applying perturbation techniques to the landmark position function, the landmark error, $\delta\mathbf{y}^e$, can be expressed as a linear function of the errors of the navigation state, terrain model, and pixel measurement model

$$\delta\mathbf{y}^e = \mathbf{G}_{yx}\delta\mathbf{x} + \mathbf{G}_{yh}\delta h + \mathbf{G}_{yz}\mathbf{v}(t_i) \tag{19}$$

where the influence coefficients

$$\mathbf{G}_{yx} = \left.\frac{\partial\mathbf{g}}{\partial\mathbf{x}}\right|_{\tilde{\mathbf{x}},\tilde{h},\tilde{\mathbf{z}}(t_i),\mathbf{C}_c^b,\mathbf{\Pi}} \tag{20}$$

$$\mathbf{G}_{yh} = \left.\frac{\partial\mathbf{g}}{\partial h}\right|_{\tilde{\mathbf{x}},\tilde{h},\tilde{\mathbf{z}}(t_i),\mathbf{C}_c^b,\mathbf{\Pi}} \tag{21}$$

$$\mathbf{G}_{yz} = \left.\frac{\partial\mathbf{g}}{\partial\mathbf{z}}\right|_{\tilde{\mathbf{x}},\tilde{h},\tilde{\mathbf{z}}(t_i),\mathbf{C}_c^b,\mathbf{\Pi}} \tag{22}$$

and

$$\delta h = \tilde{h} - h \tag{23}$$

Using the linearized position measurement, the landmark error is a zero-mean, Gaussian random vector. The landmark error covariance, $\mathbf{P}_{yy}(t_i)$, and cross-correlation matrices, $\mathbf{P}_{yx}(t_i)$, are defined as

$$\mathbf{P}_{yy}(t_i) = E[\delta\mathbf{y}\delta\mathbf{y}^T] \tag{24}$$

$$\mathbf{P}_{yx}(t_i) = E[\delta\mathbf{y}\delta\mathbf{x}^T] \tag{25}$$

Substituting (19) into (24), and noting the independence between navigation state, terrain, and pixel measurement errors yields:

$$\begin{aligned}\mathbf{P}_{yy}(t_i) &= \mathbf{G}_{yx}E[\delta\mathbf{x}\delta\mathbf{x}^T]\mathbf{G}_{yx}^T \\ &\quad +\mathbf{G}_{yh}E[\delta h^2]\mathbf{G}_{yh}^T \\ &\quad +\mathbf{G}_{yz}E[\mathbf{v}(t_i)\mathbf{v}^T(t_i)]\mathbf{G}_{yz}^T\end{aligned} \tag{26}$$

Substituting the previously defined covariance matrices for the navigation errors, terrain, and pixel measurement yields the final form of the landmark position error covariance.

$$\begin{aligned}\mathbf{P}_{yy}(t_i) &= \mathbf{G}_{yx}\mathbf{P}_{xx}(t_i)\mathbf{G}_{yx}^T + \mathbf{G}_{yh}\sigma_h^2\mathbf{G}_{yh}^T \\ &\quad +\mathbf{G}_{yz}\mathbf{R}\mathbf{G}_{yz}^T\end{aligned} \tag{27}$$

The cross-correlation matrices are calculated in a similar manner and are expressed as:

$$\mathbf{P}_{xy}(t_i) = \mathbf{P}_{xx}(t_i)\mathbf{G}_{yx}^T \tag{28}$$

$$\mathbf{P}_{yx}(t_i) = \mathbf{G}_{yx}\mathbf{P}_{xx}(t_i) \tag{29}$$

B. State Propagation

In this section, the nominal navigation state, navigation error state, and landmark error states are propagated from time t_i to t_{i+1}.

The nominal aircraft navigation state is propagated forward based on the non-linear dynamics model given in Equations (4-6), typically using a non-linear differential equation solver (e.g., Runge-Kutta) [13].

The landmark error dynamics are defined as a random walk:

$$\delta\dot{\mathbf{y}} = \mathbf{G}_y\mathbf{w}_y(t) \tag{30}$$

where $\mathbf{w}_y(t)$ is a zero-mean, white Gaussian noise process with covariance kernel

$$E[\mathbf{w}_y(t)\mathbf{w}_y^T(t+\tau)] = \mathbf{Q}_y\boldsymbol{\delta}(\tau) \tag{31}$$

The navigation error stochastic differential equation is defined in Equation (12) as

$$\delta\dot{\mathbf{x}}(t) = \mathbf{F}(t)\delta\mathbf{x}(t) + \mathbf{G}_x[\tilde{\mathbf{x}}^e(t), t]\mathbf{w}_x(t) \tag{32}$$

The navigation and landmark error covariance propagation dynamics are derived using the linearized dynamics models (12),(13),(30),(31) [9]:

$$\begin{aligned}
\dot{\mathbf{P}}_{xx}(t) &= \mathbf{F}(t)\mathbf{P}_{xx}(t) + \mathbf{P}_{xx}(t)\mathbf{F}^T(t) \\
&\quad +\mathbf{G}_x(t)\mathbf{Q}_x(t)\mathbf{G}_x^T(t) && (33)\\
\dot{\mathbf{P}}_{xy}(t) &= \mathbf{F}(t)\mathbf{P}_{xy}(t) && (34)\\
\dot{\mathbf{P}}_{yy}(t) &= \mathbf{G}_y\mathbf{Q}_y\mathbf{G}_y^T && (35)
\end{aligned}$$

An equivalent expression for the time propagation is represented by the state transition matrix, $\boldsymbol{\Phi}(t_{i+1}, t_i)$, which projects the navigation and landmark error covariance from time t_i to t_{i+1} [8]. The resulting expression for the navigation and landmark error covariance is

$$\begin{aligned}
\mathbf{P}_{xx}(t_{i+1}) &= \boldsymbol{\Phi}(t_{i+1}, t_i)\mathbf{P}_{xx}(t_i)\boldsymbol{\Phi}^T(t_{i+1}, t_i) \\
&\quad +\int_{t_i}^{t_{i+1}} \boldsymbol{\Phi}(t_{i+1}, \tau)\mathbf{G}_x\mathbf{Q}_x\mathbf{G}_x^T \cdot \\
&\quad \boldsymbol{\Phi}^T(t_{i+1}, \tau)d\tau && (36)\\
\mathbf{P}_{xy}(t_{i+1}) &= \boldsymbol{\Phi}(t_{i+1}, t_i)\mathbf{P}_{xy}(t_i) && (37)\\
\mathbf{P}_{yy}(t_{i+1}) &= \mathbf{P}_{yy}(t_i) \\
&\quad +(t_{i+1} - t_i)\mathbf{G}_y\mathbf{Q}_y\mathbf{G}_y^T && (38)
\end{aligned}$$

C. Projection of Uncertainty Statistics onto Image

The pixel projection function is used to project the navigation state and landmark location into the image plane at time t_{i+1}. The pixel projection is

$$\begin{aligned}
\mathbf{z}(t_{i+1}) = \mathbf{h}[&\mathbf{p}^e(t_{i+1}), \mathbf{C}_b^e(t_{i+1}), \mathbf{y}^e(t_{i+1}), \\
&\mathbf{C}_c^b, \boldsymbol{\Pi}]
\end{aligned} \tag{39}$$

The estimated pixel location error, $\delta\mathbf{z}(t_{i+1})$, is modeled as a perturbation about the nominal pixel location

$$\delta\mathbf{z}(t_{i+1}) = \tilde{\mathbf{z}}(t_{i+1}) - \mathbf{z}(t_{i+1}) \tag{40}$$

where the nominal pixel location, $\tilde{\mathbf{z}}(t_{i+1})$, is calculated using the nominal navigation state and landmark position

$$\begin{aligned}
\tilde{\mathbf{z}}(t_{i+1}) &= \mathbf{h}[\tilde{\mathbf{p}}^e(t_{i+1}), \tilde{\mathbf{C}}_b^e(t_{i+1}), \tilde{\mathbf{y}}^e(t_{i+1}), \\
&\quad \mathbf{C}_c^b, \boldsymbol{\Pi}]
\end{aligned} \tag{41}$$

Perturbing the pixel projection function, the pixel location error can be expressed as a linear function of the errors of the navigation state and landmark position:

$$\delta\mathbf{z}(t_{i+1}) = \mathbf{H}_{zx}\delta\mathbf{x}(t_{i+1}) + \mathbf{H}_{zy}\delta\mathbf{y}(t_{i+1}) \tag{42}$$

where

$$\begin{aligned}
\mathbf{H}_{zx} &= \left.\frac{\partial\mathbf{h}}{\partial\mathbf{x}}\right|_{\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \mathbf{C}_c^b, \boldsymbol{\Pi}} && (43)\\
\mathbf{H}_{zy} &= \left.\frac{\partial\mathbf{h}}{\partial\mathbf{y}}\right|_{\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \mathbf{C}_c^b, \boldsymbol{\Pi}} && (44)
\end{aligned}$$

The pixel error covariance, $\mathbf{P}_{zz}(t_{i+1})$, is defined as

$$\mathbf{P}_{zz}(t_{i+1}) = E[\delta\mathbf{z}\delta\mathbf{z}^T] \tag{45}$$

Substituting (42) into (45), and eliminating independent error sources yields the pixel location covariance:

$$\begin{aligned}
\mathbf{P}_{zz}(t_{i+1}) &= \mathbf{H}_{zx}\mathbf{P}_{xx}(t_{i+1})\mathbf{H}_{zx}^T \\
&\quad +\mathbf{H}_{zx}\mathbf{P}_{xy}(t_{i+1})\mathbf{H}_{zy}^T \\
&\quad +\mathbf{H}_{zy}\mathbf{P}_{xy}^T(t_{i+1})\mathbf{H}_{zx}^T \\
&\quad +\mathbf{H}_{zy}\mathbf{P}_{yy}(t_{i+1})\mathbf{H}_{zy}^T
\end{aligned} \tag{46}$$

Finally, the covariance of the pixel location errors can be summarized by combining the equations presented in the previous sections:

$$\begin{aligned}
\mathbf{P}_{zz}(t_{i+1}) = &\mathbf{H}_{zx}\boldsymbol{\Phi}(t_{i+1}, t_i)\mathbf{P}_{xx}(t_i)\boldsymbol{\Phi}^T(t_{i+1}, t_i)\mathbf{H}_{zx}^T \\
&+\mathbf{H}_{zx}\int_{t_i}^{t_{i+1}} \boldsymbol{\Phi}(t_{i+1}, \tau)\mathbf{G}_x\mathbf{Q}_x\mathbf{G}_x^T \cdot \\
&\boldsymbol{\Phi}^T(t_{i+1}, \tau)d\tau\mathbf{H}_{zx}^T \\
&+\mathbf{H}_{zx}\boldsymbol{\Phi}(t_{i+1}, t_i)\mathbf{P}_{xx}(t_i)\mathbf{G}_{yx}^T\mathbf{H}_{zy}^T \\
&+\mathbf{H}_{zy}\mathbf{G}_{yx}\mathbf{P}_{xx}(t_i)\boldsymbol{\Phi}^T(t_{i+1}, t_i)\mathbf{H}_{zx}^T \\
&+\mathbf{H}_{zy}\mathbf{G}_{yx}\mathbf{P}_{xx}(t_i)\mathbf{G}_{yx}^T\mathbf{H}_{zy}^T \\
&+\mathbf{H}_{zy}\mathbf{G}_{yh}\sigma_h^2\mathbf{G}_{yh}^T\mathbf{H}_{zy}^T \\
&+\mathbf{H}_{zy}\mathbf{G}_{yz}\mathbf{R}\mathbf{G}_{yz}^T\mathbf{H}_{zy}^T \\
&+(t_{i+1} - t_i)\mathbf{H}_{zy}\mathbf{G}_y\mathbf{Q}_y\mathbf{G}_y^T\mathbf{H}_{zy}^T
\end{aligned} \tag{47}$$

This equation shows how an initial covariance, $\mathbf{P}_{xx}(t_i)$, height uncertainty, σ_h^2, measurement noise (characterized by $\mathbf{R}$), and process noise (characterized by $\mathbf{Q}_x$ and $\mathbf{Q}_y$) can be projected to the image plane at a later time, t_{i+1}, as expressed by $\mathbf{P}_{zz}(t_{i+1})$.

In summary, given the pixel coordinates of a stationary ground landmark at time t_i, the predicted pixel coordinates of the same landmark at time t_{i+1} can be described by the bivariate Gaussian probability density function given in Equation (47). Thus, the correspondence search for the landmark can be constrained using a statistical confidence threshold. In the following section, the stochastic projection method is used to predict the location (and uncertainty) of a stationary landmark in an image.

VI. Experiment

The experiment validates the stochastic projection method using both simulated and real data collected from an airborne system. In this experiment, a Northrop T-38 "Talon" aircraft was equipped with a day-night monochrome digital video camera synchronized to a Honeywell H-764G Inertial Navigation System. The camera was mounted in the cockpit, pointing out

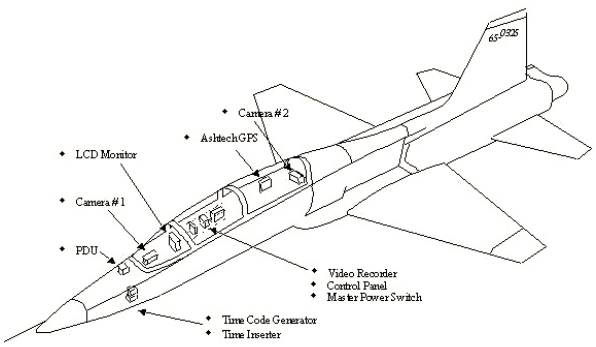

Fig. 3. Northrop T-38 instrumented with synchronized digital video camera and inertial navigation system.

the right wing. Flight data were collected in Fall of 2002 at Edwards Air Force Base, California.

A Monte Carlo simulation of the test flight is performed to verify the stochastic projection model with respect to a statistically significant sampling of random error contributors. While this provides an indication of the adequacy of the system model, flight test data are used to verify the performance of the algorithm in a real-world environment.

A. Monte Carlo Simulation

The performance of the stochastic projection method presented in Section V was verified using a statistically representative ensemble of sample functions (300 per run). The data collection system used on the T-38 flights was simulated in software, based on a reference trajectory chosen to generate an interesting observation geometry. This constant-altitude circular flight path was constructed such that a fixed terrain patch remained in the camera field of view throughout the flight. The simulated aircraft speed was 150 meters-per-second, altitude was 2296 meters, and bank angle was 27 degrees which described a circular flight path with 4592 meter radius. The resulting slant range to the landmark was 5134 meters. The terrain elevation was simulated as a zero-mean random variable. Simulations were accomplished using a terrain elevation error standard deviation of 25 meters, representing a moderate accuracy terrain model. All simulations used a 10 second interval between the first and second image, which was equivalent to 18.7 degrees of arc in the horizontal plane. The simulation geometry is shown in Fig. 4.

The results are shown in Fig. 5. In this figure, the predicted pixel location errors for each Monte Carlo sample function are represented by a "plus" symbol. The predicted 2-σ pixel location error bound is indicated by a line. Note the inclined elliptical nature of the 2-σ bound is a function of the trajectory and measurement geometry.

The same predicted pixel location errors are shown referenced to a 256×256 pixel image in Fig. 6. The stochastic constraint method shows a small correspondence search area which gives the highest probability of the landmark location. The stochastic constraint method is an improvement over the epipolar line search method as it provides a smaller search

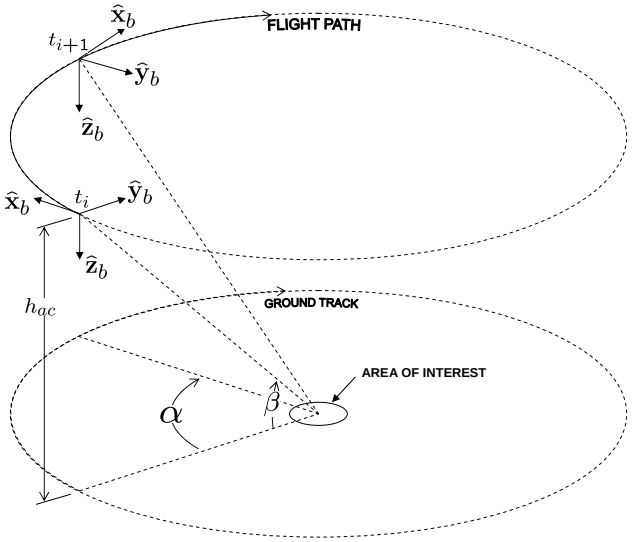

Fig. 4. Simulated flight path. In order to generate a good observation geometry, the circular orbit was chosen such that a fixed terrain patch remained in the camera field of view throughout the flight.

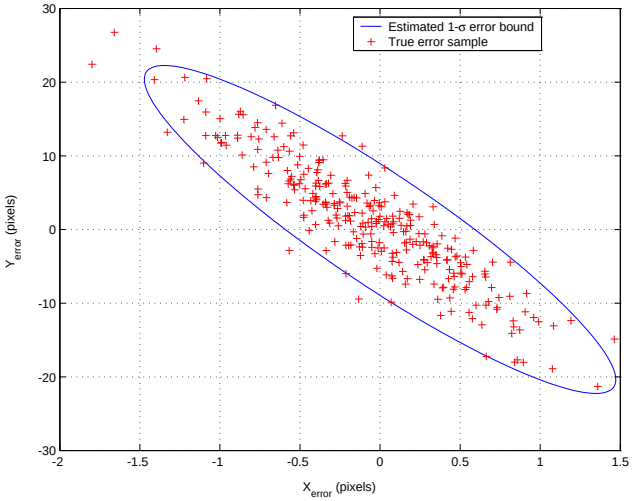

Fig. 5. Landmark pixel location error and predicted 2-σ bound for 25 meter terrain elevation uncertainty. Note the actual pixel location errors are similar to the predicted error bound. *Note: X and Y axes have differing scales to show detail.*

area developed using a statistical model. This results in faster and more robust correspondence searches.

B. Flight Data

In this section, the stochastic projection method is implemented using image and inertial flight data collected on the T-38 aircraft. The aircraft state dynamics are a function of the measurements from the strapdown inertial sensors. All states are estimated in the Earth-centered Earth-fixed reference frame previously defined. The error equations were developed based on [16], [17]. For this example, a three image sequence from a right turning profile is shown in Fig. 7. The results of the above method for predicting the future target location and uncertainty are shown in Fig. 8.

The target selected was the west corner of a building shown in Fig. 7. The estimated target location and 2-σ variance

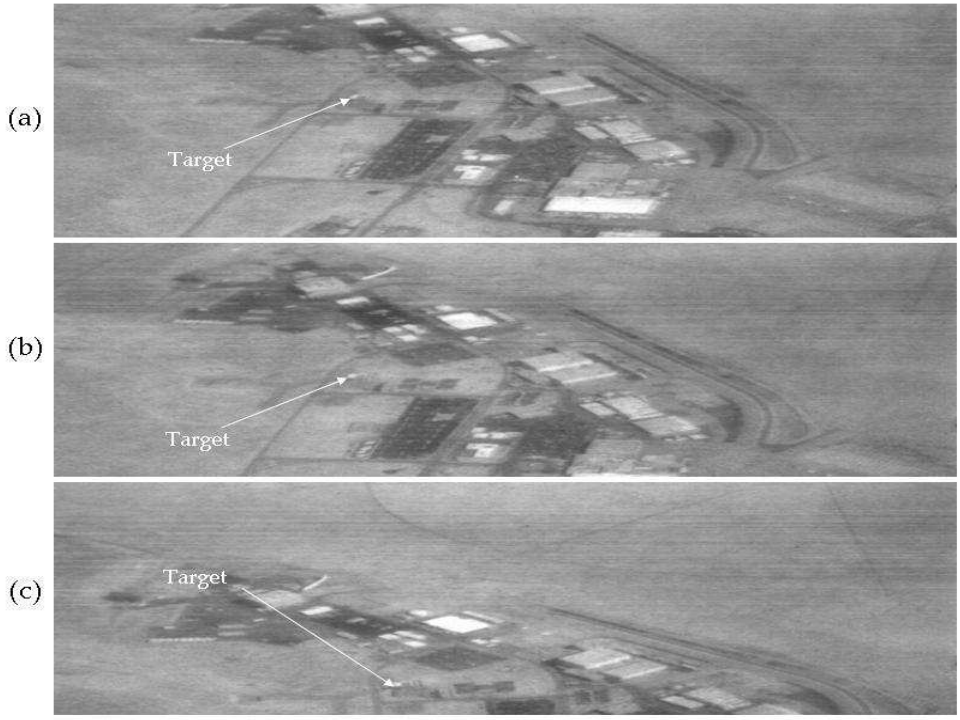

Fig. 7. Three image sequence of an industrial area recorded during a T-38 flight, with a sample stationary ground landmark identified. Image (b) was taken 1 second after image (a). Image (c) was taken 7 seconds after image (a). The aircraft is in a right turn approximately 3.8 kilometers from the landmark.

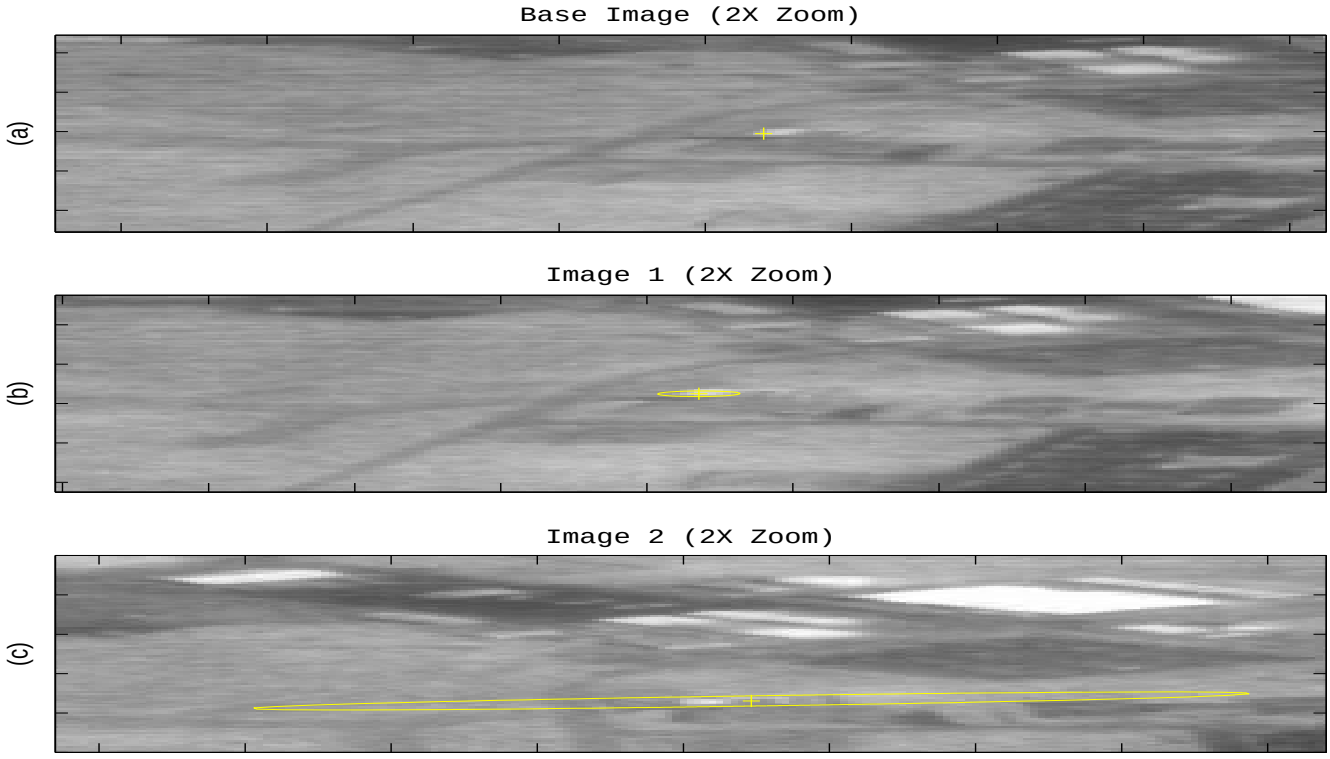

Fig. 8. Predicted landmark location uncertainty using stochastic projection method. The landmark selected was the west corner of a building in the base image (a), represented by the crosshair. Using the stochastic projection method, the landmark mean and 2-σ variance is projected into two subsequent images to demonstrate the concept. The estimated landmark location and predicted 2-σ variance for image (b) shows an ellipsoidal uncertainty after one second of flight. Image (c) shows a further increase in the uncertainty after seven seconds of flight. In each subsequent image, constraining the correspondence search for the landmark to the ellipsoidal region reduces the required search area and would eliminate false matches with other features with a similar appearance (e.g., other building corners).

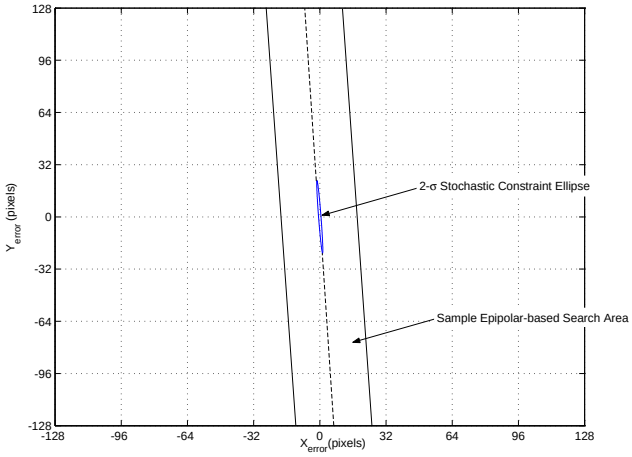

Fig. 6. Landmark pixel location error and predicted 2-σ bound for 25 meter terrain elevation uncertainty referenced to a 256×256 pixel image. Note the stochastic constraint can limit the correspondence search area significantly compared to a search near the epipolar line.

shown in Fig. 8 shows an predicted ellipsoidal uncertainty after one second and seven seconds of flight. Note the uncertainty ellipse increases with flight time, as expected. In each case, incorporating camera motion information can constrain the correspondence search space significantly. Note the true landmark location remains consistent with the predicted 2-σ uncertainty ellipse in the presence of real measurement noise and terrain model errors.

VII. Conclusions

In this paper, a stochastic projection method to incorporate the statistics of navigation dynamics and target motion models is developed to project the estimated pixel location and uncertainty of a landmark between two images. The theory is statistically rigorous. Thus, results derived from simulations and actual flight data validate the accuracy of the approach for a number of realistic scenarios.

References

[1] Henele Adams, Sanjiv Singh, and Dennis Strelow. An empirical comparison of methods for image-based motion estimation. In *Proceedings of the 2002 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems*, volume 1, pages 123–128, September 2002.

[2] Bir Bhanu, B. Roberts, and J. Ming. Inertial navigation sensor integrated motion analysis for obstacle detection. In *Proceedings IEEE International Conference on Robotics and Automation*, pages 954–959, May 1990.

[3] Espen Hagen and Eilert Heyerdahl. Navigation by optical flow. In *Proceeding of the 11th IAPR International Conference on Pattern Recognition*, volume 1, pages 700–703, 1992.

[4] Stefan Hrabar and Gaurav S. Sukhatme. A comparison of two camera configurations for optic-flow based navigation of a UAV through urban canyons. In *Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 3, pages 2673–2680, September 2004.

[5] David G. Lowe. Object recognition from local scale-invariant features. In *Proc. of the International Conference on Computer Vision*, volume 2, pages 1150–1157, September 1999. Corfu, Greece.

[6] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *Proc. DARPA Image Understanding Workshop*, pages 121–130, 1981.

[7] Yi Ma, Stefano Soatto, Jana Kosecka, and S. Shankar Sastry. *An Invitation to 3-D Vision*. Springer-Verlag, Inc., New York, New York, 2004.

[8] Peter S. Maybeck. *Stochastic Models Estimation and Control, Vol I*. Academic Press, Inc., Orlando, Florida 32887, 1979.

[9] Peter S. Maybeck. *Stochastic Models Estimation and Control, Vol II*. Academic Press, Inc., Orlando, Florida 32887, 1979.

[10] Clark F. Olson, Larry H. Matthies, Marcel Schoppers, and Mark W. Maimone. Robust stereo ego-motion for long distance navigation. In *Proceedings of the IEEE Conference on Advanced Robotics*, volume 2, pages 453–458, June 2000.

[11] Meir Pachter and Alec Porter. Bearings-only measurements for INS aiding: The three-dimensional case. In *Proceedings of the 2003 AIAA Guidance, Navigation and Control Conference*, 2003. AIAA paper number 2003-5354.

[12] Meir Pachter and Alec Porter. INS aiding by tracking an unknown ground object - theory. In *Proceedings of the American Control Conference*, volume 2, pages 1260–1265, 2003.

[13] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipies in C++ Second Edition*. Cambridge University Press, The Edinburgh Building, Cambridge CB2 2RU, United Kingdom, 2002.

[14] John F. Raquet and Michael Giebner. Navigation using optical measurements of objects at unknown locations. In *Proceedings of the 59th Annual Meeting of the Institute of Navigation*, pages 282–290, June 2003.

[15] Dennis W. Strelow. *Motion Estimation from Image and Inertial Measurements*. PhD thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, November 2004.

[16] D.H. Titterton and J.L. Weston. *Strapdown Inertial Navigation Technology*. Peter Peregrinus Ltd., Lavenham, United Kingdom, 1997.

[17] M. Wei and K. P. Schwarz. A strapdown inertial algorithm using an Earth-fixed cartesian frame. *Journal of the Institute of Navigation*, 37(2):153–167, Summer 1990.